

Kevin Elkin

elkinkm97@gmail.com ❖ (720) 849-1148 ❖ San Diego, CA

EXPERIENCE

Verisk 3E

NLP Data Scientist

San Diego, CA

Feb 2021 – Present

One of five data scientists on the R&D and Automation team building the company SaaS product using ML and NLP to automate the extraction of key information from technical chemical documents for customers of Verisk 3E.

- Designed and implemented a language agnostic text extraction solution that utilized deep learning (NER + BERT) and GPT-3 to extract desired technical information to an underlying database and automated 73% of previous manual extraction efforts.
- Built an ensemble model using Tesseract OCR that prescreened candidate documents resulting in a higher success rate and saving 25% of the annual AWS Textract spend.
- Presented applicable use cases of NLP, LLMs, and ChatGPT across companywide town hall meetings to both technical and non-technical audiences to increase the adoption across additional projects.
- Secured buy in from senior management for a new product feature proposal to use cloud and ML services to analyze chemical documents and derive a green index rating based on ecological impact.
- Migrated projects to the cloud (AWS) and retrained legacy ML models that made use of the following services: EC2, S3, lambda, SageMaker, AWS Glue, and OpenSearch.

Qualcomm

Data Scientist Intern

San Diego, CA

Jun 2019 – Sep 2019

- Automated 2 monthly dashboards and reports on WAN Optimization and Robotics Process Automation (RPA) using Tableau and Cisco's Technology Business Management model (TBM) for the Office of the CIO.
- Utilized Tableau to create various metrics, data visualizations, and dashboards and directly pulled data from Splunk in real time.
- Communicated 25+ data visualizations, metrics, and designs across multiple teams within IT through weekly meetings to ensure the dashboard(s) fit the client's needs; awarded a QualStar for outstanding work and achievement.

Allied Universal

LiveSafe Database Administrator

Denver, CO

Dec 2018 – Jun 2019

- Monitored multiple databases to track all 15,000+ tenants and company personnel in San Francisco, Denver, and Dallas.
- Redesigned and implement innovative ways to monitor and inform personnel by categorizing users through 22 custom geofences.

Brookfield Properties

Central Operations Officer

Denver, CO

Jun 2017 – Dec 2018

- Oversaw 20+ computer systems in a control setting environment and communicated to security and local authorities via radio.

EDUCATION

University of California, San Diego

Bachelor of Science in Data Science, Minor in Business

La Jolla, CA

Sep 2016 – Dec 2020

- **Major GPA:** 3.73
- **Relevant Coursework:** Data Analysis and Inference, Data Structures, Web Mining and Recommender Systems, Systems for Scalable Analytics, Data Visualization, Data Management, Probability and Statistics in Data Science, Artificial Intelligence: Probabilistic Reasoning and Decision-Making, Theoretical Foundations of Data Science, Linear Algebra, Machine Learning

PROJECTS AND RESEARCH

HinDroid Malware Classification (Link: <https://github.com/KevinElkin/Hindroid>)

Jan 2020 – Apr 2020

- Created a data ingestion pipeline that extracts, transforms, and loads (ETL) benign and malicious Android applications from the apkpure.com (sitemap) and decompiles the applications into smali code.
- Utilized graph embeddings to create a heterogeneous information network (HIN) to classify benign and malware Android apps.
- Calculated kernels to represent relationships between different API calls that encompasses invoke type, package type, code block, and applications.
- Trained and tested a LinearSVC model on a dataset containing 1,000 benign and 1,000 malware apps (80-20 split) and assessed results based on an F1 score of .981, precision, recall, accuracy, and a confusion matrix.

SKILLS AND INTERESTS

- **Languages/Technologies:** Proficient in Python, SQL, Pandas, NumPy, scikit-learn, LangChain, GPT, ChatGPT, BeautifulSoup, NLTK, Regex, Keras, PyTorch, AWS (EC2, lambda, S3, SageMaker, Glue, Comprehend, Textract), Java, Dask, Spark, Docker, Tableau, MS Excel, MATLAB, GitHub, Azure Dev Ops; familiar with R, HTML, CSS, JavaScript, D3.